Human Activity Identification Using Pose Estimation

Neha Marne,Akansha Jagtap,Isshita Paliwal

Sinhgad College Of Engineering, Vadgao,Pune,Maharashtra,India

Abstract

Human activity recognition plays a significant role in human-to-human/machine interaction and interpersonal relations. The machine's ability to recognize another person's activities is one of the main subjects of study of the scientific areas of computer vision and machine learning. With the advent of Pose estimation along with Classification algorithms, which canbe used on images/video input key point detection and prediction, it is now possible to collect and store data on different aspects of human movement under the conditions of free living and identify the action performed.This technology has the potential to be used in automated activity profiling systems which produce a continuous record of activity patterns over extended periods of time. Such activity profiling systems are dependent on classification algorithms which can effectively interpret motion data and identify different activities. This paper reviews the pose estimation technique(mediapipe) which have been used to classify normal activities and/or identify falls from body-joints data. The review is structured according to the different analytical techniques and illustrates the variety of approaches which have previously been applied in this field. Although significant progress has been made in this important area, there is still significant scope for further work, particularly in the application of advanced classification techniques to problems involving many different activities.

Keywords: - pose estimation, mediapipe, classification, human activity recognition, key points

Introduction

Human activity recognition has a wide range of uses because of its impact on wellbeing. Human activities have an inherent hierarchical structure that indicates the different levels of it, which can be considered as a three-level categorization. Movements are often typical activities performed indoors, such as walking, talking, standing, and sitting. They may also be more focused activities such as those types of activities performed in a kitchen or on a factory floor. Human Activity Recognition is the problem of predicting what a person is doing based on a trace of their movement using Pose estimation and Classification algorithm. It is a challenging problem because there are many motions involved to specific actions in a general way. Human activity recognition is a field of study that deals with identifying, interpreting, and analysing the actions specific to the movement of human beings. This project will analyse the activity being performed by the user in the Video/Image input. Human activity recognition will use Pose estimation and classification algorithm to analyse the data set and detect the activity. Human activity recognition basis for many applications such as video surveillance, health care, and human-computer interaction. To analyse the activity of a person from the information collected by different devices. Discover which are the variables that determine the activity. To Calculate a predictive model that can recognize a person's activity from the signals received by the sensors. Design individualized exercise tables to improve the health of a person.

Proposed System

Pose estimation - Pose Estimation is a computer vision task that infers the pose of a person or object in an image or video. Pose estimation is the problem of determining the position and orientation of a camera relative to a given person or object. This is typically done by identifying, locating, and tracking a number of keypoints on a given object or person. For

objects, this could be corners or other significant features. And for humans, these keypoints represent major joints like an elbow or knee. The goal of our machine learning models are to track these keypoints in images and videos.

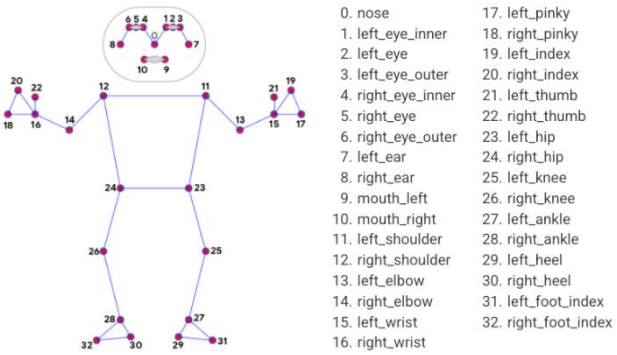

Fig 1.Body keypoints

Dataset Creation - Based on the mediapipe analysis we create our CSV file to train our Classification Algorithm and Prediction is made about the human activity. The admin can use webcam for live input or upload video input and according to the details the prediction will be made. Logistic Regression was used to predict the action.

	class	x1	y1	z1	v1
0	Standing	0.441929	0.179151	-0.768026	0.999997
1	Standing	0.441217	0.179151	-0.949387	0.999997
2	Standing	0.440749	0.179168	-0.946997	0.999997
3	Standing	0.440405	0.179345	-0.940684	0.999997
4	Standing	0.440183	0.179372	-0.959358	0.999996

Fig 2.Dataset format

Class - action information/name (6 classes)

x and y - Landmark coordinates normalized to [0.0, 1.0] by the image width and height respectively.

z: Represents the landmark depth with the depth at the midpoint of hips being the origin, and the smaller the value the closer the landmark is to the camera. The magnitude of `z` uses roughly the same scale as `x`.

visibility: A value in [0.0, 1.0] indicating the likelihood of the landmark being visible (present and not occluded) in the image.

Namaste	468
Standing	424
Hands_On_Waist	420
Bend_Down	331
Left_Hand_Up	272
Right_Hand_Up	252

Fig 3.Actions in dataset

Classification Algorithm

We define the activity classification problem as a multiple class classification problem, which can be modelled using different machine learning classification and regression algorithms. The classification algorithm takes 33 body keypoints dataset (x-axis, y-axis z-axis and visibility values of each point) as input for the model's training and testing. We used a supervised learning approach as we have labelled dataset containing body keypoints with an activity label. Among all the algorithms, we use logistic regression which provides 100% accuracy.

Logistic Regression - It is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no). Mathematically, a logistic regression model predicts P(Y=1) as a function of X. Logistic regression models the data using the sigmoid function.

$$f(x) = \frac{1}{1+e^{-(x)}}$$

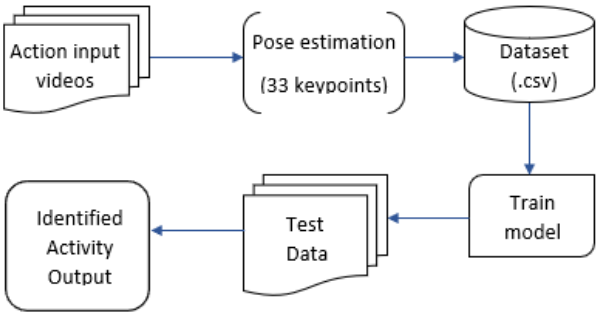

Fig 4.System Workflow

Evaluation Metrics

Precision Precision(P) – It is the ratio of the number of true positives (Tp) to the sum of false positives (Fp) and true positives. It can also be defined as how many images classified into this class belong to this class.

$$P = Tp/(Tp+Fp)$$

Recall (R) - It is the ratio of the number of true positives (Tp) to the sum of false negatives (Fn) and true positives. It can also be defined as how many images that belong to this class are classified into this class.

$$R = Tp/(Tp+Fn)$$

F1-Score – It is calculated as the harmonic mean of recall and precision. Equation calculates it.

$$F1\text{-}Score = 2\ (P*R)/(P+R)$$

Confusion Matrix - It is a two-dimensional matrix used to measure the overall performance of the machine learning classification algorithm. In the matrix, each row is associated with the predicted activity class, and each column is associated with the actual activity class. The matrix compares the target activity with the activity predicted by the model. This gives a better idea of what types of errors our classifier has made.

Results and Discussion

We have used Logistic regression, ridge classifier, random forest classifier, gradient boosting classifier and got 100% accuracy for Logistic regression, random forest classifier, gradient boosting classifier and 99% for ridge classifier. We used logistic regression as our final model for prediction.

Algorithm	Precision	Recall	F1-Score	Support
Logistic regression	1.0	1.0	1.0	651
Ridge classifier	0.99	0.99	0.99	651
Random forest classifier	1.0	1.0	1.0	651
Gradient boosting classifier	1.0	1.0	1.0	651

Fig 5. Evaluation of algorithms

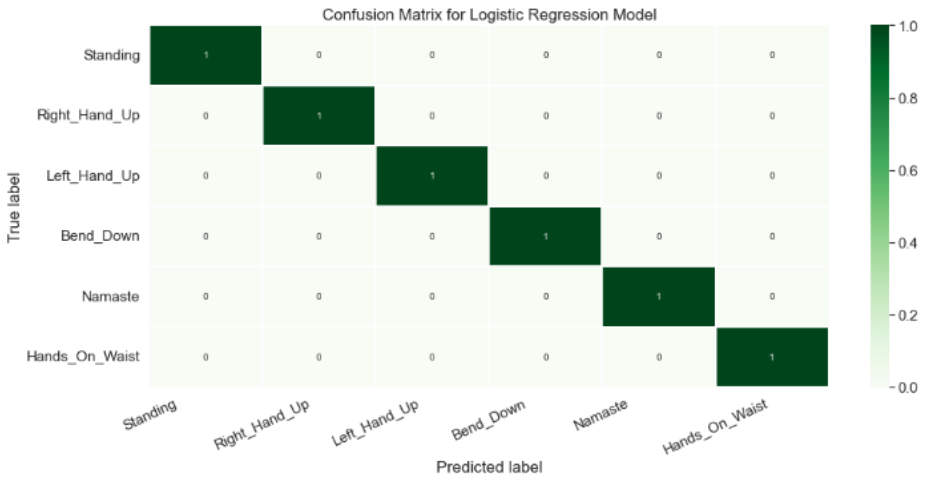

Fig 6.Confusion Matrix for Logistic Regression

Conclusion

In this project we have introduced Human Activity Identification technique which will detect the activity being performed in the input video. We have proposed a Human Activity Identification system based on pose estimation and Classification algorithm. This system will combine the results of the 3D pose estimation model with the classification technique for better and more accurate detailed action prediction. We have used Logistic regression, ridge classifier, random forest classifier, gradient boosting classifier and got 100% accuracy for Logistic regression, random forest classifier, gradient boosting classifier and 99% for ridge classifier. We used logistic regression as our final model for prediction.

References

[1] Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. Computer vision and image understanding, 73(3), 428-440.

[2] Kamel, A., Sheng, B., Yang, P., Li, P., Shen, R., & Feng, D. D. (2018). Deep convolutional neural networks for human action recognition using depth maps and postures. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(9), 1806-1819.

[3] Hossain, M. R. I., & Little, J. J. (2018). Exploiting temporal information for 3d human pose estimation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 68-84).

[4] H. P. S. A. Lateef and U. Eranna, "Modelling Approach to Select Potential Joints for Enhanced Action Identification of Human," 2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2018, pp. 1506-1509, doi: 10.1109/ICEECCOT43722.2018.9001313.

[5] R. A., Amiruzzaman, M., Ahmed, N., & Islam, M. R. (2020, August). Efficient Frequency Domain Feature Extraction Model using EPS and LDA for Human Activity Recognition. In 202020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII) (pp. 344-347). IEEE.

[6] Gupta, Abhay, Kuldeep Gupta, Kshama Gupta, and Kapil Gupta. "Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm."